

На правах рукописи

Синдеев Михаил Сергеевич

**Исследование и разработка алгоритмов матирования
видеопоследовательности**

Специальность 05.13.11 – математическое и программное обеспечение
вычислительных машин, комплексов и компьютерных сетей

Автореферат

диссертации на соискание учёной степени
кандидата физико-математических наук

Москва – 2013

Работа выполнена в Институте прикладной математики им. М.В.Келдыша РАН.

Научный руководитель:

кандидат физико-математических наук, доцент Баяковский Юрий Матвеевич, доцент, ф-т ВМиК Московского государственного университета им. М.В.Ломоносова

Официальные оппоненты:

Соколов Сергей Михайлович. Профессор, д. ф.-м. н., зав. сектором. Институт прикладной математики им. М. В. Келдыша РАН.

Сафонов Илья Владимирович. Доцент, к. т. н., Московский инженерно-физический институт.

Ведущая организация:

Московский государственный технический университет им. Н. Э. Баумана

Защита состоится « 21 » мая 2013 г. в 11 часов на заседании диссертационного совета Д 002.024.01, созданного на базе Института прикладной математики имени М.В.Келдыша РАН по адресу: 125047, Москва, Миусская пл., 4.

С диссертацией можно ознакомиться в библиотеке Института прикладной математики им. М.В.Келдыша РАН.

Автореферат разослан «___» апреля 2013 г.

Учёный секретарь диссертационного совета,
д.ф.-м.н.

Полилова Т.А.

Общая характеристика работы

Объект исследования и актуальность работы

В работе рассматривается задача матирования – выделения объектов в изображении или видеопоследовательности с целью монтажа, т.е. последующего наложения объекта на новый фон. Решение задачи матирования заключается в вычислении маски прозрачности, называемой «альфа-каналом», и цвета каждого пиксела объекта. Критерием качества матирования является незаметность монтажа для зрителя.

Данная задача возникла в художественной фотографии довольно давно, и долгое время решалась трудоемкими аналоговыми методами. Широкое применение матирование нашло в кино: неподвижные или подвижные (покадровые) маски («маты») объекта переднего плана рисовались на стеклянных панелях и предотвращали экспонирование фона на пленку, куда затем отдельным проходом экспонировался новый фон на основе инвертированной маски. Сведение слоев могло также осуществляться оптическими способами, например, проецированием через полупрозрачное зеркало.

Из-за сложностей в создании покадровых масок кинематографисты часто ограничивались неподвижными масками, заведомо захватывающими область перемещения объектов переднего плана (чаще всего – актеров), при этом фон за ними брался из реальной сцены, а дорисовка фона был ограничена областью, в которую объекты переднего плана заведомо не попадали.

Матирование занимает важное место в профессиональной обработке видео и кинопроизводстве и применяется для замены/модификации фона, цветокоррекции отдельных объектов, а также для преобразования видео в стереоскопический (3D) формат.

Помимо визуальных эффектов в видео, потенциальной областью применения является дополненная реальность. Существующие технологии позволяют дополнять видеопоток синтетическими объектами в реальном времени, однако возможности по бесшовному совмещению этих объектов с реальными ограничены – они, как правило, просто накладываются на входное изображение, в то время как желательно обрабатывать перекрытия искусственных объектов реальными.

На рис. 1 показан пример кадра из видеопоследовательности, альфа-канал (маска прозрачности) и результат наложения на новый фон.

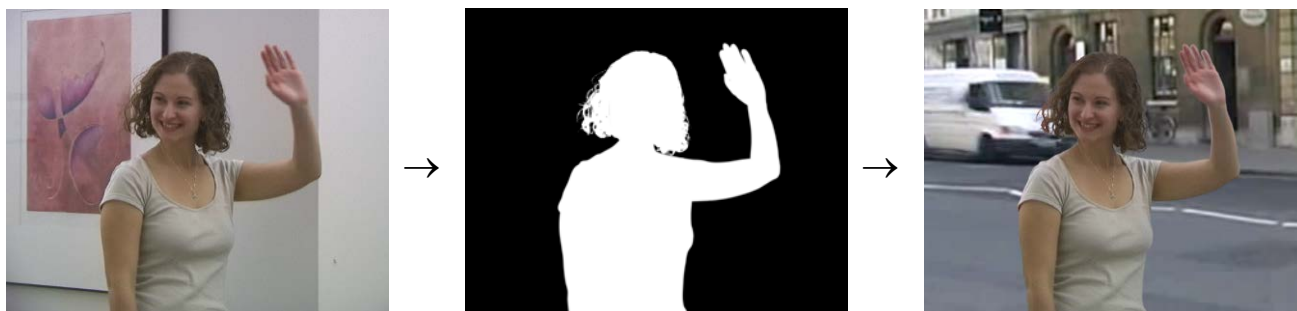


Рисунок 1. Пример матирования кадра из видеопоследовательности

В данной работе рассматриваются подходы к матированию изображений и видеопоследовательностей. Второй случай является более сложным, так как требует согласованности масок в соседних кадрах, чтобы избежать эффекта мерцания. Естественным подходом к обеспечению такой согласованности является использование оптического потока, представляющего собой карту межкадрового движения (векторное поле скоростей для каждого пиксела). В работе предложена идея вычисления оптического потока в альфа-канале, что делает поток пригодным для задачи матирования, устраняя некоторые недостатки обычного оптического потока.

Во введении формулируются цели и задачи работы, рассматриваются ограничения на исходные данные, обосновывается актуальность и показывается практическая значимость работы. В первой главе рассматривается задача матирования изображений. Во второй главе формулируется задача матирования видеопоследовательностей по ключевым кадрам и предлагается алгоритм для ее решения. В третьей главе рассматривается задача вычисления оптического потока (и, в качестве частного случая, альфа-потока). Предложены два алгоритма – более быстрый двухкадровый и более точный траекторный. В четвертой главе описывается алгоритм матирования видеообъема. Предложено разбиение видеообъема на временные суперпикселы.

Цель диссертационной работы

Целью работы является исследование существующих подходов к задаче отделения объекта переднего плана от фона в видеопоследовательности, анализ их недостатков, формализация задачи, определение требований к входным данным; разработка алгоритма матирования видеопоследовательностей, верификация разработанного алгоритма путем оценки качества его работы на реальных данных и сравнение предложенного метода с существующими аналогами; создание программных модулей для матирования видео на основе разработанного алгоритма.

Полученная система должна превосходить существующие по соотношению качество результата / объем пользовательского ввода. Основным критерий качества результата – при наложении извлеченного слоя на новый фон монтаж должен быть не заметен. Также задачей этой работы является разработка формального количественного критерия для оценки результатов.

Научная новизна работы

В рамках диссертации разработаны новые алгоритмы матирования изображений и видеопоследовательностей и вычисления оптического потока. Отличительной чертой предложенного алгоритма матирования является межкадровая согласованность канала прозрачности и повышенная устойчивость к ошибкам вычисления оптического потока.

Практическая значимость и реализация

В рамках работы были реализованы программные модули, которые могут быть использованы как независимо, так и в комплексе. Алгоритм матирования

изображений на основе байесовского подхода реализован в виде подключаемого модуля «GrowCut 3.0» к программе Adobe Photoshop. Данный модуль позволяет в интерактивном режиме строить и уточнять канал прозрачности для выделения объекта на изображении. Также реализованы модули вычисления оптического потока, сегментации и матирования видеопоследовательности.

На основе данных модулей реализована программная система для матирования видеопоследовательности на языках Matlab и C++. Предложенный алгоритм разрабатывался в рамках проекта с компанией «Microsoft Research Cambridge».

Апробация работы

Результаты работы докладывались и обсуждались на

- 17-ой международной конференции по компьютерной графике и машинному зрению «Graphicon'2007», Россия, Москва, 2007
- 18-ой международной конференции по компьютерной графике и машинному зрению «Graphicon'2008», Россия, Москва, 2008
- 19-ой международной конференции по компьютерной графике и машинному зрению «Graphicon'2009», Россия, Москва, 2009
- 16-й международной научной конференции студентов, аспирантов и молодых учёных «Ломоносов-2009»
- 5-й летней школе Microsoft для аспирантов (Microsoft Research PhD Summer School), Великобритания, г. Кембридж, 2010
- семинаре группы компьютерного зрения Microsoft Research, 23 июля 2010), Великобритания, г. Кембридж, 2010
- 22-ой международной конференции по компьютерной графике и машинному зрению «Graphicon'2012», Россия, Москва, 2012
- Семинаре по компьютерной графике и машинному зрению Ю.М. Баяковско-го (ф-т ВМК МГУ)
- Семинаре направления «Программирование» им. М. Р. Шура-Бура в ИПМ им. М. В. Келдыша РАН

Публикации

По теме диссертации автором опубликовано 7 научных работ, в т.ч. 2 в журналах ВАК [6], [7]. Статья, посвященная предложенному алгоритму матирования видеопоследовательностей, была принята на ведущую международную конференцию ACCV-2012 и опубликована в журнале Lecture Notes in Computer Science издательства Springer [7].

Структура и объем работы

Диссертация состоит из введения, четырех глав, заключения и списка литературы. Содержание работы изложено на 116 страницах. Список литературы включает 84 наименования. В работе содержится 43 рисунка и 3 таблицы.

Содержание работы

Во введении описывается задача матирования изображений и видео. Обоснована актуальность диссертационной работы, сформулированы цели и задачи, аргументирована научная новизна исследований, показана практическая значимость полученных результатов. Описана структура диссертации.

Рассмотрены ограничения на исходные данные. Предполагается, что на вход алгоритма подается видеопоследовательность длиной примерно до 10 секунд в цветовом формате RGB. В ней должен присутствовать явный объект переднего плана, т.е. предполагается, что сцена снята оператором (человеком), который может контролировать процесс съемки с учетом того, что затем потребуется выделять объект из видео. Объект должен быть виден на протяжении всей видеопоследовательности. В кадре не должно быть резких движений, изменений освещенности и резкой смены видимого изображения объекта. Разрешение до 1280x720, частота кадров – 24-25. Обрабатываемый видеофрагмент должен быть доступен целиком, т.е. работа в реальном времени не осуществляется.



Рисунок 2. Уравнение смешивания.

При соблюдении этих требований достигаются наилучшие результаты, продемонстрированные в данной работе. На практике допустимы ослабления требований, но тогда может потребоваться создание дополнительных ключевых кадров.

Для видеопоследовательностей с объектом переднего плана справедлива модель формирования изображения на основе уравнения смешивания $C = \alpha F + (1 - \alpha) B$. Формулируется задача восстановления канала прозрачности α и изображений F , B , если известно только изображение C . Уравнение смешивания проиллюстрировано на рис. 2.

Задача является недоопределенной, поэтому все существующие алгоритмы опираются на пользовательский ввод или дополнительные данные (например, многокурсовые фотографии). Так как в общем случае такие данные недо-

ступны, большинство алгоритмов полагается на пользовательский ввод. Появление новых алгоритмов обусловлено стремлением уменьшить время, затрачиваемое пользователем на разметку данных.

Перечислены и проиллюстрированы основные проблемы существующих методов, из-за которых применение этих методов на практике затруднено и требует значительного количества пользовательского ввода для исправления ошибок. Таким образом, целью работы является создание алгоритма, лишенного этих недостатков.

Предложена общая схема алгоритма, использующего два ключевых кадра (в начале и в конце видеопоследовательности) в виде граничных условий на искомый канал прозрачности, как проиллюстрировано на рис. 3.

Также введено понятие оптического потока и описаны существующие алгоритмы матирования видео, сгруппированные по разным критериям, в том числе по способу использования оптического потока. Проанализированы проблемы существующих методов, которые также сгруппированы по типам.



Рисунок 3. Схематичное изображение видеопоследовательности с двумя ключевыми кадрами

В первой главе рассматривается задача матирования изображений. Это требуется для матирования ключевых кадров, а также как основа алгоритма матирования видеопоследовательности. Формулируется понятие тернарной разметки как формата пользовательского ввода для уточнения задачи (см. рис. 4).

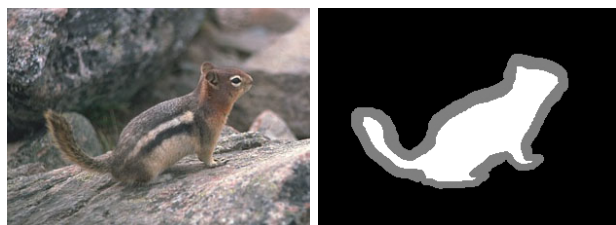


Рисунок 4. Пример тернарной разметки для изображения.

Затем рассматривается предложенный автором модифицированный алгоритм матирования, основанный на байесовском подходе¹:

$$P(F, B, \alpha | C) = \frac{P(C | F, B, \alpha)P(F)P(B)P(\alpha)}{P(C)}$$

$P(\alpha)$ моделируется нормальным распределением с мат. ожиданием, вычисленным по соседним уже обработанным пикселям и дисперсией, вычисленной как мера близости распределений $P(F)$ и $P(B)$. Для вычисления мат. ожидания $P(\alpha)$ предложена сортировка пикселей по цветовой близости с уже обработанными пикселями.

Модифицированный алгоритм дает более устойчивый результат в случаях, когда цветовая информация неоднозначна (распределения для F и B близки или совпадают) за счет использования предположения о пространственной когерентности значений прозрачности. Также повышается связность результата – устраняются «дырки» и «островки» в маске объекта.

Для ускорения вычислений применен иерархический подход с использованием идеи метода аналитического матирования² для ускорения вычислений.

Предложен алгоритм автоматической генерации входной тернарной разметки на основе бинарной, полученной алгоритмом *GrowCut*³. Такой подход уменьшает время, требуемое пользователю для создания разметки. Кроме того, реализована возможность прямого редактирования маски прозрачности с автоматическим пересчетом значений F и B , что экономит время в случаях, когда проще уточнить непосредственно значение прозрачности, чем подгонять разметку для достижения такого же уточнения.

Проведена количественная оценка⁴ точности предложенного метода и продемонстрированы преимущества предложенного метода по сравнению с исходным вариантом байесовского алгоритма. Описана реализация алгоритма в виде подключаемого модуля к программе Adobe Photoshop.

Во второй главе формулируется задача матирования видеопоследовательности по ключевым кадрам и предлагается алгоритм для ее решения с использованием оптического потока в альфа-канале (альфа-потока). Данный алгоритм опирается на два других алгоритма – вычисления оптического потока и матирования видеообъема, которые описаны в последующих главах. Данные алгоритмы не являются фиксированными: предложенный метод может быть применен с использованием почти любых алгоритмов, решающих данные подзадачи, что позволит улучшить результаты матирования при появлении эффективных алгоритмов для этих подзадач в будущем.

¹ Chuang Y., Curless B., Salesin D., Szeliski R. A Bayesian Approach to Digital Matting // In proc. of CVPR. 2001. P. 264–271

² Levin A., Lischinski D., Weiss Y. A Closed Form Solution to Natural Image Matting // In proc. of CVPR. 2006. P. 61–68

³ Vezhnevets V., Konouchine V. Grow-Cut – Interactive Multi-Label N-D Image Segmentation by Cellular Automata // In proc. of Graphicon. 2005. P. 150–156

⁴ Тестовая база: Rhemann C., Rother C., Wang J., Gelautz M., Kohli P., Rott P. A Perceptually Motivated Online Benchmark for Image Matting // In proc. of CVPR. 2009. P. 1826–1833. <http://www.alphamatting.com/>

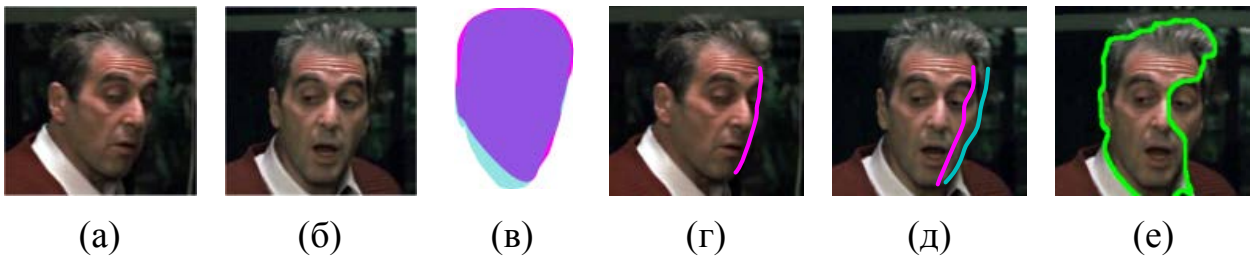


Рисунок 5. Иллюстрация идеи алгоритма: (а, б) первый и последний кадры фрагмента, (в) маски, соответствующие этим кадрам, (г) контур в первом кадре, (д) два способа переноса контура на последний кадр – в предположении, что контур лежит на поверхности, и в предположении, что контур является видимой границей объекта, (е) результат алгоритма SnapCut⁵ – видно, что данный алгоритм предполагает принадлежность контура поверхности, что неверно.

Основной идеей алгоритма является вычисление оптического потока в канале прозрачности. Такой подход позволяет совместить маску с видимым контуром объекта. Идея предложенного метода вытекает из наблюдения, что при перекрытиях форма видимого контура, в отличие от изображения объекта, меняется не сильно. Данный принцип проиллюстрирован на рис. 5.

Оптический поток в канале прозрачности является регуляризацией для α , а не условием связи изображений и прозрачности, как сделано в существующих методах. Такая регуляризация позволяет использовать произвольное число полных и частичных ключевых кадров. Связь прозрачности с изображениями при этом обеспечивается двумя способами:

- использованием обычного оптического потока с небольшим весом
- использованием алгоритма матирования изображений, обобщенного на видеопоследовательность с учетом перекрытий (этот подход изложен в четвертой главе)

Схема алгоритма приведена на рис. 6. Алгоритм использует обычный оптический поток в цветовых каналах RGB в качестве начального приближения. Затем после матирования он уточняется за счет использования альфа-канала. RGB каналы также используются в качестве регуляризации, но с меньшим весом.

В главе предлагается следующая формулировка целевого функционала задачи матирования:

$$E(V, \alpha, M) = \iiint \left[\left(\frac{\partial \alpha}{\partial V} \cdot M \right)^2 + \nu J(I, \alpha) + \mu \left(\frac{\partial I}{\partial V} \cdot M \right)^2 + \lambda \|\nabla V\|^2 + \rho(M, I, V) \right] dx dy dt,$$

где M – бинарная маска перекрытий,

ρ – условие разреженности маски перекрытий,

J – целевой функционал матирования для каждого кадра,

⁵ Bai X., Wang J., Simons D., Sapiro G. Video SnapCut: robust video object cutout using localized classifiers // In proc. of SIGGRAPH. 2009. P. 1–11

$\frac{\partial}{\partial V}$ – производная вдоль вектора потока V .

Решение задачи матирования видеопоследовательности сводится к минимизации данного функционала с граничным условием на значения прозрачности в ключевых кадрах. Минимизация производится поочередно по переменным V , M , α . Слагаемые функционала энергии описаны подробнее в следующих главах.

Преимуществом предлагаемого алгоритма является временная согласованности результата (альфа-канала), достигаемая за счет условия гладкости потока в этом канале, а не в оптическом потоке (который является промежуточным представлением и влияет на результат косвенно).

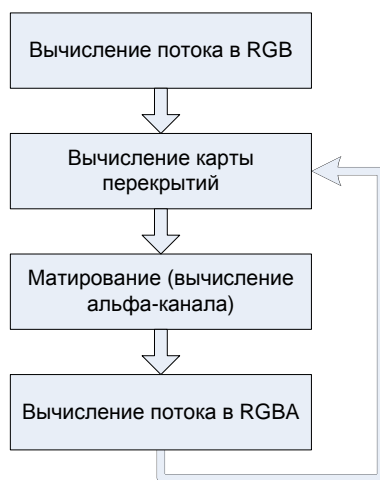


Рисунок 6. Схема предложенного алгоритма матирования видео

Предложенный алгоритм матирования видеопоследовательности является итерационным и многократно вызывает алгоритмы вычисления потока и матирования для всех кадров. Отсюда вытекает необходимость сделать данные алгоритмы быстрыми, чему посвящены две следующие главы.

В третьей главе рассматривается задача вычисления оптического потока как в цветовых каналах, так и в канале прозрачности.

Оптическим потоком называется векторное поле видимого движения пикселей между кадрами. Задача его поиска формулируется так: по данной паре изображений I_1 , I_2 построить оптический поток (u, v) , совмещающий эти изображения:

$$I_1(x, y) \approx I_2(x + u(x, y), y + v(x, y)).$$

Предложены два новых алгоритма – более быстрый двухкадровый и более точный траекторный. За основу взята вариационная формулировка задачи оптического потока, заключающаяся в минимизации функционала энергии⁶

⁶ Alvarez L., Esclarin J., Lefebure M., Sanchez J., A PDE model for computing the optical flow // In proc. of XVI Congreso de Ecuaciones Diferenciales y Aplicaciones. 1999. P. 1349–1356

$$E(V) = \iint \left[\left(\frac{\partial I}{\partial V} \right)^2 + \lambda \|\nabla V\|^2 \right] dx dy.$$

Помимо непрерывных методов также существуют дискретные, основанные на переборе вектора сдвига (u, v) по множеству $\{-M, \dots, M\}^2$, где M – максимальный сдвиг.

Предложенный двухкадровый метод является комбинацией непрерывного и дискретного подходов и использует метод отдельной оптимизации для поиска минимума функционала энергии. В данном методе неизвестное поле V дополняется аналогичным полем U и слагаемые энергии применяются к этим полям независимо. Также добавляется мягкое условие равенства U и V :

$$E(U, V) = \iint \left[\left(\frac{\partial I}{\partial V} \right)^2 + \theta(U - V)^2 + \lambda \|\nabla U\|^2 \right] dx dy.$$

Устремляя параметр θ к бесконечности, получим минимум данной энергии, в котором $U = V$, и достигается совместный минимум слагаемых данных и гладкости. Минимизация производится итерационно с увеличением параметра θ . Каждое из слагаемых само по себе эффективно минимизируется предназначенным для него алгоритмом.

Слагаемое данных минимизируется переборным алгоритмом на основе алгоритма PatchMatch⁷. Вместо фрагментов изображения используются единичные пиксели, т.к. за пространственную когерентность отвечает слагаемое гладкости потока. Основная идея метода – оптимизация процедуры перебора гипотез векторов оптического потока. Пусть область поиска имеет размер $M \times M$. Тогда вместо перебора M^2 значений в каждом пикселе осуществляется выбор одного случайного значения в каждом пикселе, а затем используется двухпроходное распространение решения на соседние пиксели. Ввиду пространственной когерентности (обусловленной как структурой естественных изображений, так и непосредственно слагаемым гладкости) многие связанные области пикселей имеют почти одинаковые значения потока, поэтому если оптимальный вектор найден хотя бы в одном пикселе области, он распространится на всю область. Шаблон обработки пикселей показан на рис. 7.

Поток в канале прозрачности получается заменой I на α :

$$E_\alpha(V) = \iint \left[\left(\frac{\partial \alpha}{\partial V} \right)^2 + \lambda \|\nabla V\|^2 \right] dx dy.$$

В качестве дополнительной регуляризации, которая важна на первых итерациях, когда значения α не достаточно точны, используется поток для каналов RGBA, т.е. энергия является суммой энергий оптического и альфа- потоков:

$$E_{RGB\alpha}(V) = \iint \left[w_{RGB} \left(\frac{\partial I}{\partial V} \right)^2 + \left(\frac{\partial \alpha}{\partial V} \right)^2 + \lambda \|\nabla V\|^2 \right] dx dy,$$

Где w_{RGB} – вес цветочных каналов.

⁷ Barnes C., Shechtman E., Finkelstein A., Goldman D. PatchMatch: A randomized correspondence algorithm for structural image editing // Communications of the ACM. 2011. V 54, N 11. P. 103–110

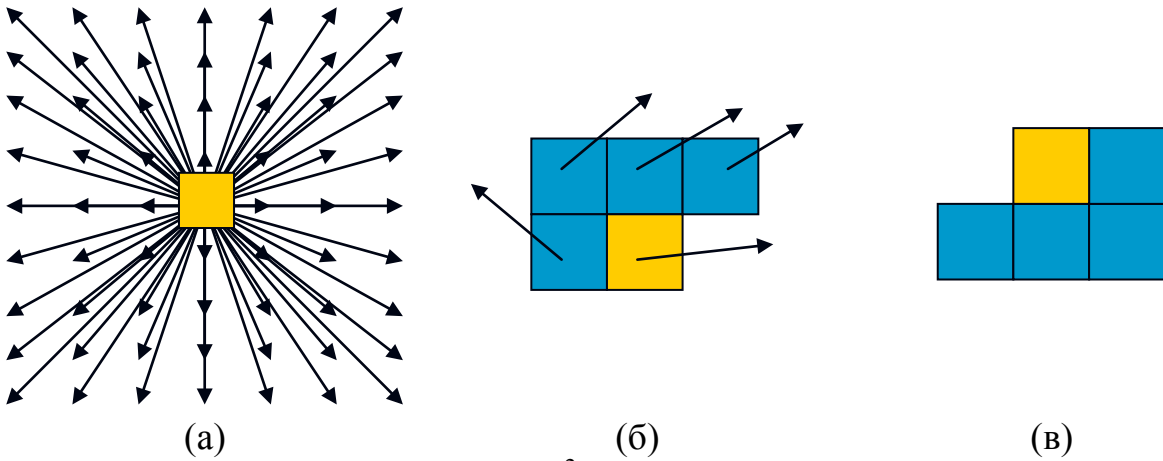


Рисунок 7. (а) Полный перебор (M^2 возможных векторов в каждом пикселе). (б) Перебор методом PatchMatch – вектор потока выбирается из 5 кандидатов при прямом проходе (текущий пиксел и 4 ранее обработанных пиксела). (в) шаблон обратного прохода.

Второй предложенный алгоритм – траекторный поток – является более точным, чем первый, но более медленным. Основная идея алгоритма заключается в использовании нескольких кадров сразу для нахождения перекрытий и уточнения слагаемого данных. В данном алгоритме рассматривается $T = 2K + 1$ кадров с номерами $-K, \dots, K$. Финальным результатом является оптический поток из кадра I_0 в I_1 и обратный поток из I_0 в I_{-1} , но при этом в качестве промежуточного результата находятся $2K$ потоков из I_0 в остальные кадры. В каждом пикселе кадра I_0 эти потоки образуют траекторию.

Т.к. потоки вычисляются относительно кадра I_0 , их можно считать одним «траекторным» потоком, где каждому пикселу сопоставлен $4K$ -мерный (или, что то же самое, $2(T - 1)$ -мерный) вектор потока

$$\vec{V} = (u_{-K}, v_{-K}, \dots, u_{-1}, v_{-1}, u_1, v_1, \dots, u_K, v_K).$$

Такое векторное представление позволяет использовать любые известные функционалы гладкости, определенные для двумерного потока, которые могут быть сформулированы в терминах векторной алгебры, а также многие алгоритмы минимизации данных функционалов. Исключением будут алгоритмы, явно использующие факт двумерности, а также алгоритмы, несовместимые с членом данных E_D (однако их часто можно адаптировать для данной задачи, либо использовать методы раздельной оптимизации, используя другой алгоритм для члена данных). Кроме того, стохастические и переборные алгоритмы могут потерять свою эффективность из-за увеличения пространства поиска. Метод траекторного потока использует следующий функционал энергии:

$$E(V, \rho) = E_D(V, \rho) + \lambda E_S(V) + \mu E_T(V) + \nu E_\rho(\rho),$$

в котором слагаемое данных определено с учетом видимости:

$$E_D(V, \rho) = \frac{1}{K} \sum_x \sum_y \sum_{\substack{\rho(x,y) \\ t \neq 0}}^{\rho(x,y)-K} (I_0(x, y) - I_t(x + u_t, y + v_t))^2,$$

слагаемое пространственной гладкости уравнивает разброс значений по разным кадрам, путем деления на $|t|$:

$$E_s(V) = \frac{1}{T-1} \sum_x \sum_y \left[\sum_{\substack{t=-K \\ t \neq 0}}^K \left(\frac{\nabla V}{|t|} \right)^2 \right]^\gamma,$$

при этом норма градиента является инвариантной к повороту при $\gamma=0,5$, что в двухмерном случае записывается как

$$\psi(V) = \sqrt{\left(\frac{\partial u}{\partial x} \right)^2 + \left(\frac{\partial u}{\partial y} \right)^2 + \left(\frac{\partial v}{\partial x} \right)^2 + \left(\frac{\partial v}{\partial y} \right)^2},$$

причем в суммирование по t также производится под знаком радикала. Это означает, что в случае разрывности решения разрыв потока в некоторой точке происходит одновременно в горизонтальной и вертикальной компонентах потока и сразу во всех кадрах. Таким образом, траектории в двух соседних пикселях не могут разойтись, а через несколько кадров слиться в единый объект.

Энергия временной гладкости использует вторую производную потока по времени, т.е. поощряет траектории, близкие к линейным:

$$E_T(V) = \frac{1}{T-2} \sum_x \sum_y \sum_{t=-K+1}^{K-1} (V_{t-1} - 2V_t + V_{t+1})^2.$$

На маску видимости также наложено условие гладкости:

$$E_p(\rho) = \sum_x \sum_y (\nabla \rho(x, y))^2,$$

где градиент является конечной разностью, т.к. значения ρ дискретны. На рис. 8 показан пример траекторий для одномерного случая. Видимые части траекторий не перекрываются с соседними (хотя явного условия, запрещающего такие перекрытия нет), что говорит о корректности работы алгоритма (т.е. слагаемые гладкости и карта видимости правильно распознают траектории, относящиеся к разным движущимся слоям).

Для минимизации функционала энергии используется метод глобальной оптимизации QRVO⁸ со структурой графа, показанной на рис. 9. Данный метод является бинарным и «склеивает» решения-кандидаты. В работе предложено несколько эвристик для генерации такого решения. Часть из них заключается в небольшом возмущении текущего решения (например, сдвиг на 1 пиксел по оси x), другая часть осуществляет локальную оптимизацию неполного функционала энергии (включающего не все слагаемые).

В частности, предложено применение метода PatchMatch непосредственно к траекториям. Также предложено усовершенствование этого метода: вместо выбора наилучшего кандидата \vec{V} по шаблону из 5 пикселей предлагается строить новое решение, вообще говоря не совпадающее с этими 5 кандидатами (но совпадающее с одним из них в каждом кадре). Для этого используется алгоритм динамического программирования. Метод проиллюстрирован на рис. 10.

⁸ C. Rother, V. Kolmogorov, V. Lempitsky, M. Szummer. Optimizing binary MRFs via extended roof duality. CVPR, 2007.

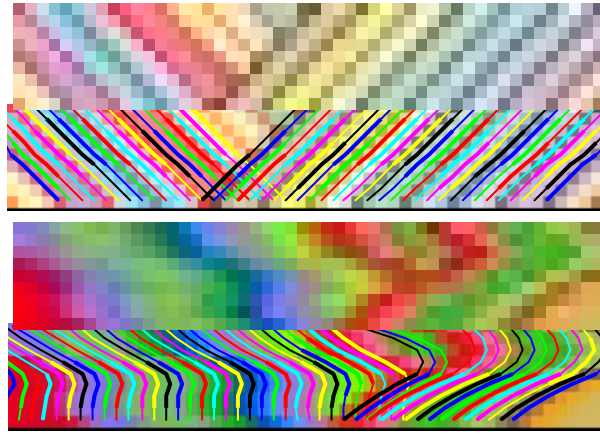


Рисунок 8. Искусственные одномерные примеры траекторного потока. Показаны исходные одномерные видеопоследовательности ($T = 9$ кадров, по вертикали – ось времени) и найденные траектории. Толстые части траекторий – диапазоны видимости.

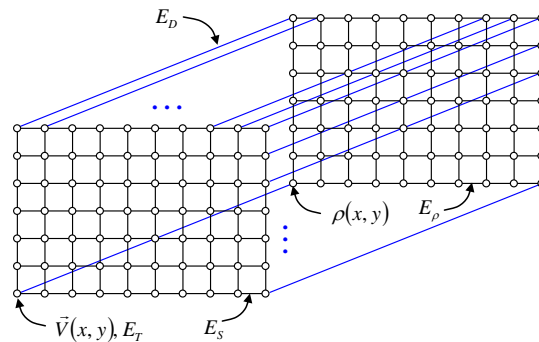


Рисунок 9. Структура графа для совместной минимизации V и ρ методом QPBO.

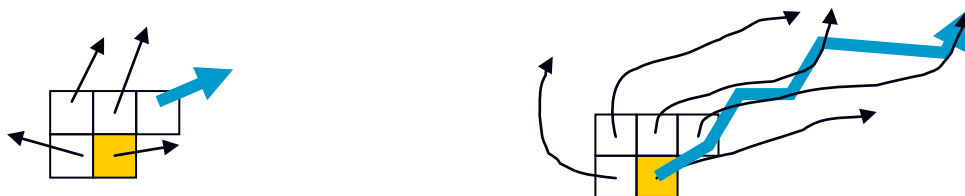


Рисунок 10. Обобщение алгоритма PatchMatch для траекторий: (а) оптический поток – выбор наилучшего вектора из 5 кандидатов; (б) траекторный поток – синтез новой траектории на основе 5 кандидатов.

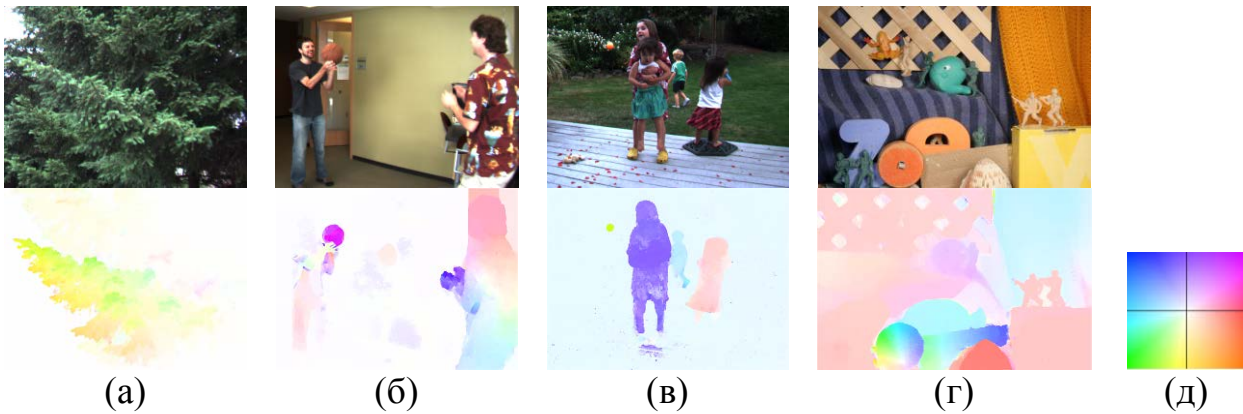


Рисунок 11. Результаты на тестовой базе Middlebury. (а-г) изображения и полученный оптический поток, (д) цветовой ключ визуализации потока.

Экспериментальная оценка метода проводилась на тестовой базе Middlebury⁹. Примеры приведены на рис. 11.

В четвертой главе описывается алгоритм матирования видеообъема. За основу взят алгоритм аналитического матирования изображения и предложено его обобщение на произвольный граф. Предложено разбиение видеообъема на временные суперпиксели как способ построения такого графа по видеопоследовательности и оптическому потоку (см. рис. 12). Данный подход учитывает возможные перекрытия, т.е. неявным образом задает функционал $\rho(M, I, V)$ в энергии $E(V, \alpha, M)$.

В процессе матирования предполагается, что значение прозрачности не меняется внутри суперпикселя. Такая модель является наиболее удобной и экономит время и память, т.к. число неизвестных равно числу суперпикселей. В работе также проведено сравнение с полной моделью, в которой значения прозрачности попиксельны.

Для построения суперпикселей предложен жадный алгоритм минимизации энтропии. Каждый суперпиксел хранит:

- средний цвет μ , дисперсию σ^2
- начало t_1 , длительность τ .

Каждый пиксел суперпикселя хранит:

- цвет относительно μ
- координату $x(t), y(t)$.

Тогда, используя принцип минимальной длины описания, можно потребовать от разбиения видеообъема на суперпиксели минимальность их общей энтропии

$$H = H_{\mu} + \tau \cdot H_p + \tau \cdot H_c,$$

где H_{μ} – число битов для хранения среднего цвета и дисперсии (константа),

⁹ S. Baker, D. Scharstein, J. Lewis, S. Roth, M. Black, R. Szeliski. A Database and Evaluation Methodology for Optical Flow. International Journal of Computer Vision. Volume 92, Number 1. 2011.

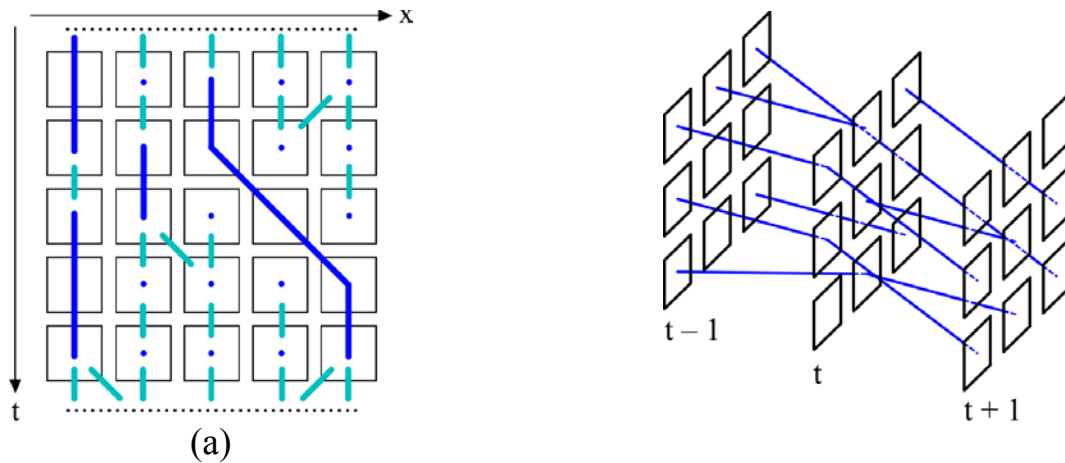
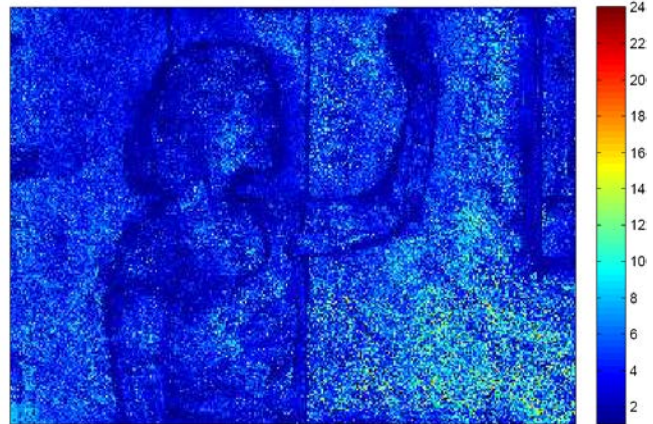


Рисунок 12. Построение временных суперпикселов. (а) Одномерное видео. Голубым показаны неиспользованные векторы потока, синим – суперпикселы (в т.ч. вырожденные), (б) иллюстрация случая двухмерного видео. Лапласиан видеообъема строится на основе показанных связей и двумерных связей соседних пикселов в каждом кадре.



(а)



(б)

Рисунок 13. Иллюстрация разбиения на суперпикселы: (а) кадр видеопоследовательности, (б) длины траекторий (в кадрах), проходящих через пикселы этого кадра.

H_p – число битов для кодирования координат одного пиксела ($x(t)$, $y(t)$) – также константа,

H_c – число битов для хранения отклонения цвета пиксела от среднего цвета μ в предположении, что это отклонение подчиняется нормальному распределению с дисперсией σ^2 .

Энтропия нормального распределения определяется как

$$H_c(\sigma) = \frac{1}{2} \ln(2\pi e \sigma^2).$$

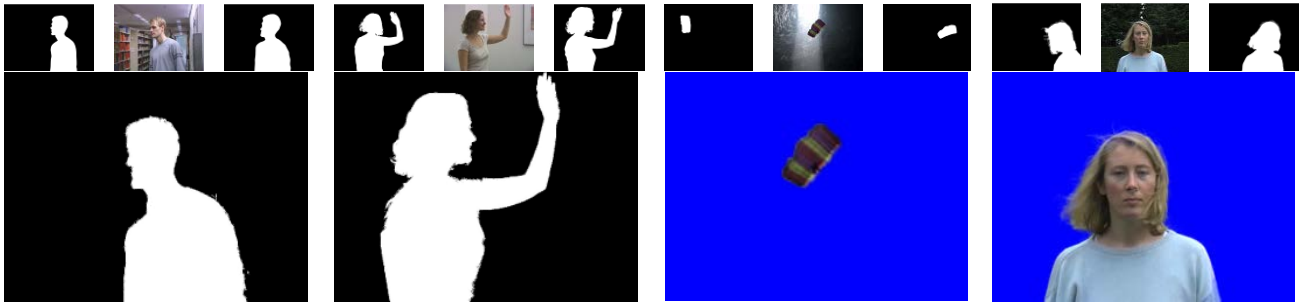


Рисунок 14. Результаты работы алгоритма. Вверху: ключевые кадры и промежуточный кадр. Внизу: результат для промежуточного кадра.

Жадный алгоритм использует двоичную кучу для хранения векторов потока и соответствующих им значений уменьшения энтропии. После выбора вектора, дающего наибольшее уменьшение, суперпикселы, соединяемые этим вектором объединяются (обновляются значения μ , σ) и пересчитываются значения уменьшения энтропии для векторов, инцидентных новому (объединенному) суперпикселу.

На рис. 13 проиллюстрировано разбиение видеообъема на суперпикселы. Видно, что на границе движущегося объекта суперпикселы преимущественно вырожденные (состоят из одиночных пикселов), а в областях без перекрытий более длинные, что обеспечивает связность видеообъема. За счет этого финальный результат матирования содержит меньше артефактов. Примеры результатов работы алгоритма приведены на рис. 14.

За счет совместного использования предложенного метода поиска оптического потока, суперпикселов для поиска перекрытий, а также ключевых кадров, можно обрабатывать видеопоследовательности с межкадровым движением около 15-20 пикселов (при разрешении 640x480), что в 3-4 раза превосходит соответствующую оценку для метода Лукаса-Канаде¹⁰, равную 5 пикселов на кадр.

Для количественной оценки результатов предложена метрика, оценивающая временную стабильность результата (отсутствие межкадрового дрожания):

$$A(t) = \sum_{x,y} \alpha(x, y, t) \quad \text{– площадь объекта,}$$

$$A^*(t) = \sum_{x,y} \alpha^*(x, y, t) \quad \text{– площадь объекта в эталонной разметке,}$$

$$\tilde{A}(t) = kA(t) + b \quad \text{– нормализованная площадь объекта путем оптимизации}$$

¹⁰ Lucas B., Kanade T. An iterative image registration technique with an application to stereo vision // Proceedings of Imaging Understanding Workshop. 1981. P. 121 – 130

$$\sum_t (\tilde{A}(t) - A^*(t))^2 \rightarrow \min \quad \text{по } k, b$$

$$\frac{1}{T} \int_0^T \left| \frac{d\tilde{A}}{dt} - \frac{dA^*}{dt} \right| dt \quad - \text{ финальная метрика стабильности.}$$

Нормализация площади объекта гарантирует инвариантность метрики к неточностям в эталонной разметке (наиболее типичная ситуация – маска объекта морфологически расширена или сжата на 1-3 пиксела).

Предложенный алгоритм превосходит существующие как по данной метрике, так и по суммарной квадратичной ошибке (таблица 1). Также продемонстрирована устойчивость алгоритма по входным данным (ключевым кадрам). При использовании подхода на основе двух ключевых кадров сохраняется целостность результирующей маски при внесении незначительных изменений в ключевые кадры, в то время как алгоритм Video SnapCut выдает некорректный результат..

Таблица 1. Численное сравнение алгоритмов

Алгоритм	Оценка дрожания	Квадратичная ошибка
Предложенный алгоритм	232,9	1418,7
Алгоритм отслеживания [Tsai 2010]	481,5	2068,1
Video SnapCut	614,9	8201,1
Объемный алгоритм CFS	1051,7	25612,7
Video Object Cut&Paste	1719,4	14060,3

В **заключении** перечислены основные результаты работы.

Основные результаты работы

В результате работы разработаны новые алгоритмы и программные модули, позволяющие построить канал прозрачности для объекта в видеопоследовательности при меньшем объеме входных данных по сравнению с существующими аналогами. Модули матирования изображений и вычисления оптического потока также могут быть использованы независимо для других задач.

1. На основе проведенного исследования методов матирования изображений и видеопоследовательностей предложен новый алгоритм матирования видеопоследовательностей, позволяющий обеспечить временную стабильность выделения движущихся объектов, используя оптический поток в канале прозрачности.
2. Разработан новый алгоритм матирования изображений на основе байесовского подхода для создания ключевых кадров, превосходящий стандартный байесовский алгоритм по качеству результата.

3. Разработаны и реализованы новые алгоритмы вычисления оптического потока, позволяющие ускорить и повысить качество матирования видео.
4. На базе разработанных алгоритмов реализованы программные модули для вычисления оптического потока и извлечения объектов из изображений и видеопоследовательностей, позволяющие ускорить и автоматизировать процесс фотомонтажа в реальных практических приложениях.
5. Разработана метрика и реализован модуль для оценки качества матирования.

Публикации по теме диссертации

1. Sindeyev M., Konushin V., Vezhnevets V. Improvements of Bayesian Matting // Proc. of Graphicon. 2007, P. 88–95
2. Sindeyev M., Konushin V. A Novel Interactive Image Matting Framework // Proc. of Graphicon. 2008, P. 41–44
3. Sindeyev M., Konushin V. A Novel Approach to Video Matting using Optical Flow // Proc. of GraphiCon. 2009, P. 340–343
4. Синдеев М., Конушин В. Матирование видео на основе оптического потока // Сборник тезисов XVI Международной научной конференции студентов, аспирантов и молодых учёных «Ломоносов-2009», секция «Вычислительная математика и кибернетика». 2009. С. 75
5. Синдеев М., Конушин А., Ротер К. Многокадровый оптический поток на основе траекторий. Труды конференции Графикон. 2012. С. 288–291
6. Синдеев М., Конушин В. Интерактивное байесовское матирование изображений // Программные продукты и системы. 2012. N 4. С. 167–171
7. Sindeev M., Rother C., Konushin A. Alpha-Flow for Video Matting // Asian Conference on Computer Vision (ACCV) 2012, Part III, Lecture Notes in Computer Science, 2013, Volume 7726, P. 438–452

Личный вклад автора

Все результаты, составляющие основное содержание диссертации, получены автором самостоятельно. В работах [1, 2, 6] автору принадлежат идеи условия гладкости прозрачности в байесовской модели, упорядоченной обработки пикселей по цветовой близости, иерархической обработки, инструмента построения тернарной разметки на основе алгоритма Дейкстры и инструментов редактирования канала прозрачности с пересчетом цветочных каналов. В работах [3, 4] автором предложен новый алгоритм совместного вычисления оптического потока и канала прозрачности по одному ключевому кадру. В работе [5] автором предложен новый алгоритм вычисления оптического потока. В работе [7] автором предложен новый алгоритм совместного вычисления оптического потока и канала прозрачности по двум ключевым кадрам, а также алгоритм сегментации видеопоследовательности на временные суперпиксели и алгоритм быстрого вычисления оптического потока на основе раздельной оптимизации.